Large Language Models (in Sweden, the Nordics and Europe)

AI
S W E D E N

magnus.sahlgren@ai.se

Language models learn a probability distribution over language

Pretraining + finetuning (+ alignment)

General text processing capabilities
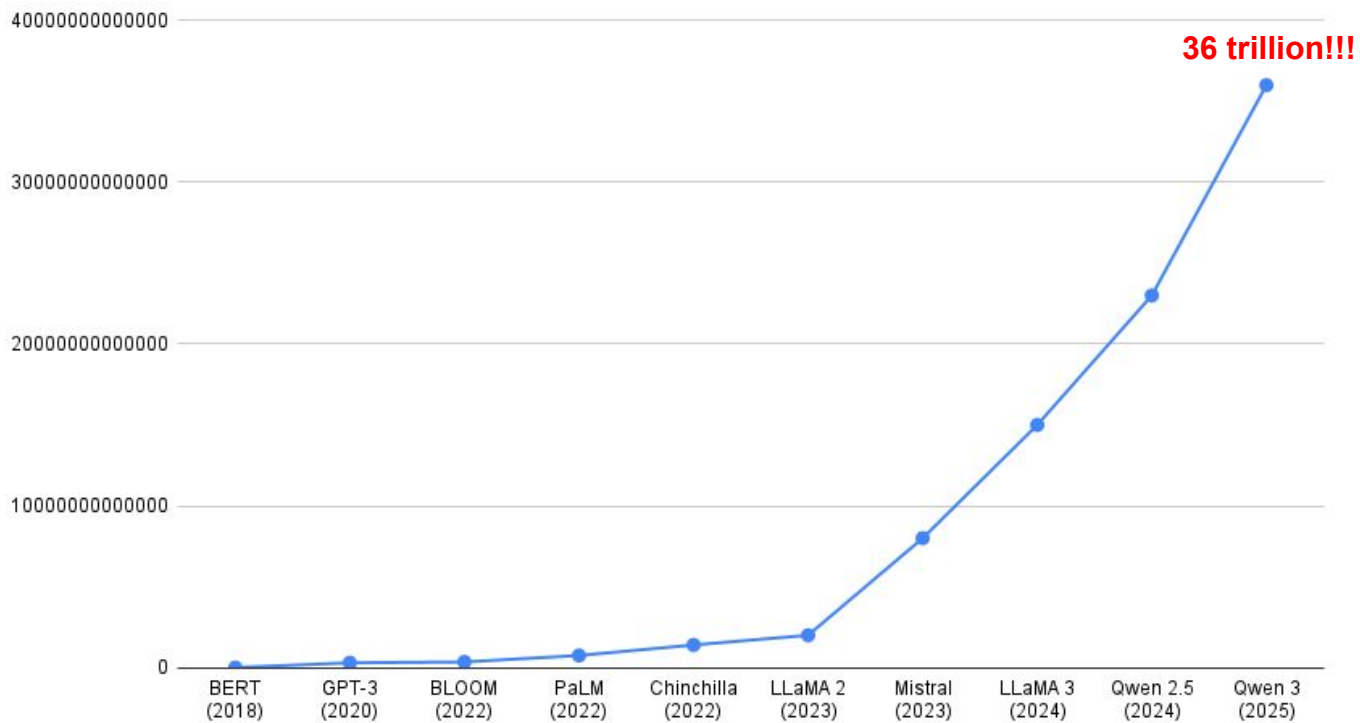
Size of training data

Size of training data

36 trillion!!!

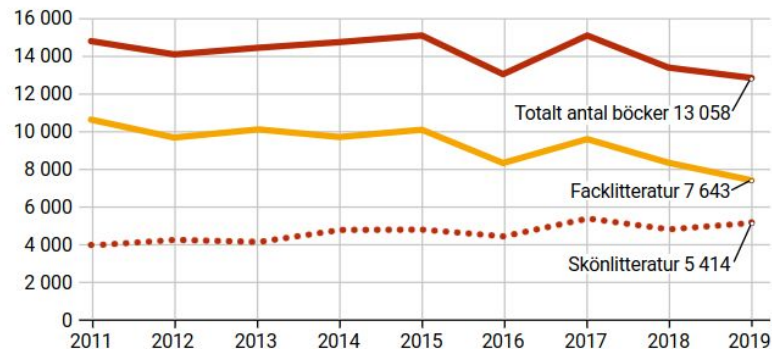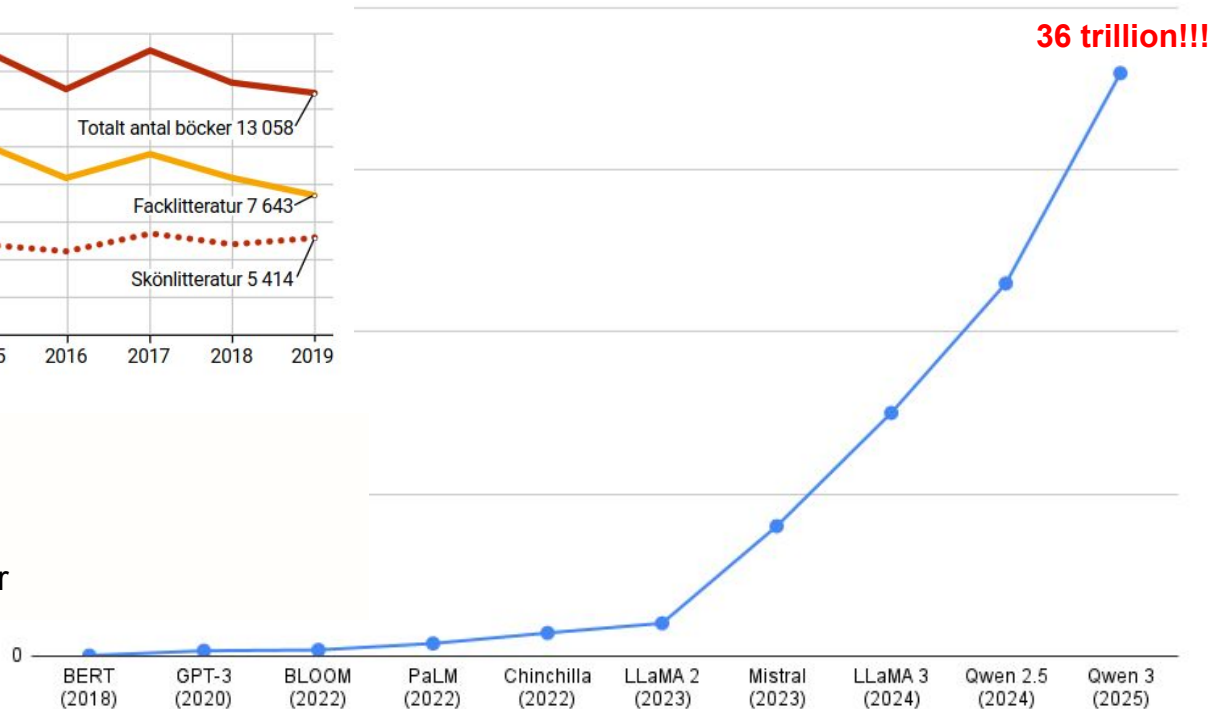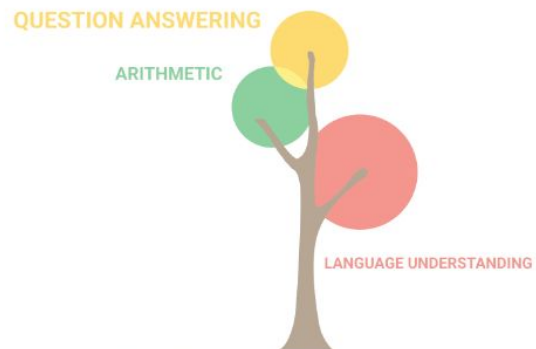X-axis: BERT (2018), GPT-3 (2020), BLOOM (2022), PaLM (2022), Chinchilla (2022), LLaMA 2 (2023), Mistral (2023), LLaMA 3 (2024), Qwen 2.5 (2024), Qwen 3 (2025)

Y-axis: 0, 10000000000000, 20000000000000, 30000000000000, 40000000000000

magnus.sahlgren@ai.se
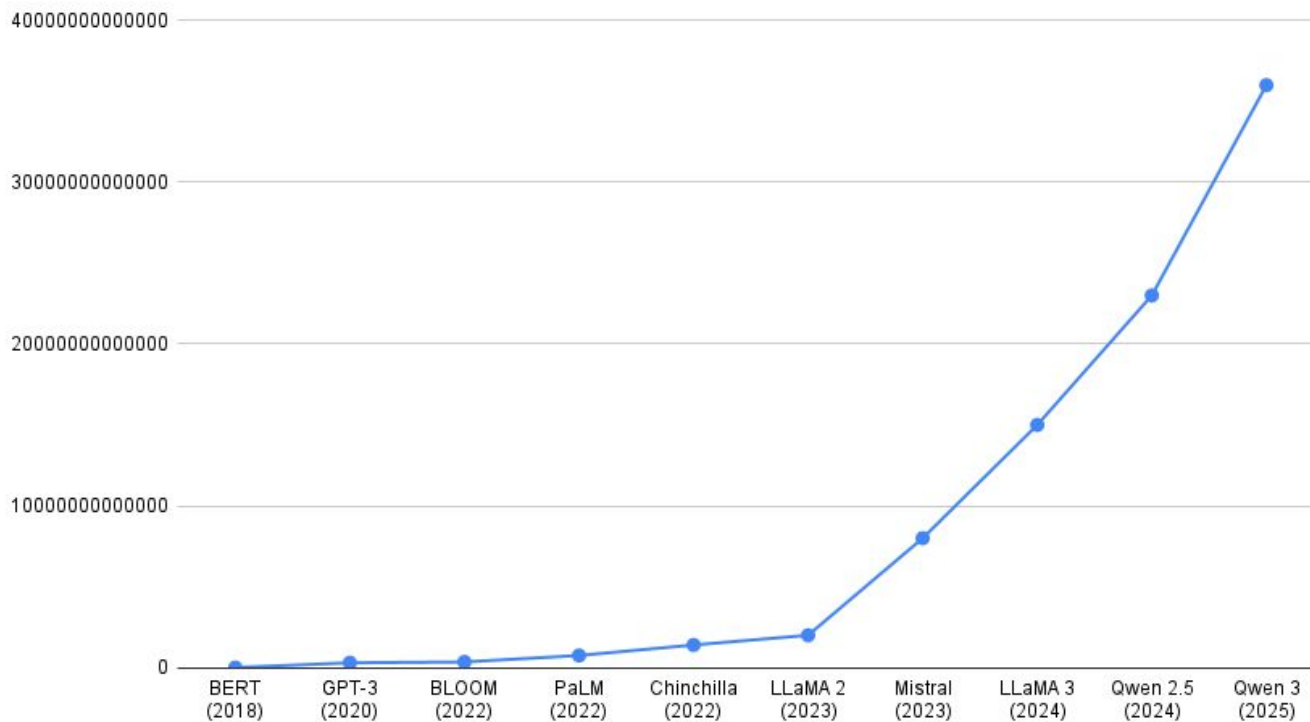
Diagram 4 Totalt antal böcker i nationalbibliografin 2011–2019

Totalt antal böcker 13 058
Facklitteratur 7 643
Skönlitteratur 5 414

36 trillion!!!

300 pages per book,
300 words per page
≈
1,25 billion words per year

BERT (2018), GPT-3 (2020), BLOOM (2022), PaLM (2022), Chinchilla (2022), LLaMA 2 (2023), Mistral (2023), LLaMA 3 (2024), Qwen 2.5 (2024), Qwen 3 (2025)

magnus.sahlgren@ai.se

QUESTION ANSWERING

ARITHMETIC

LANGUAGE UNDERSTANDING

8 billion parameters

ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html

Language models in Sweden

2017: Transformer introduced by Google

2018: the first Transformer-based language models (GPT, BERT)

2019: language models for Swedish authorities (Vinnova)

2020: the first LLM (GPT-3)

2021: NLU-group at AI Sweden (Vinnova)

2022: the first Scandinavian LLM: GPT-SW3 (Vinnova)

2023: GPT-SW3 as open model

2024: AI Sweden + EU (EuroLingua-GPT, TrustLLM, DeployAI…)

2025: Open models in EU (OpenEuroLLM)

AI
S W E D E N

GPT-SW3 consortium

Data

Infrastructure

Competence

AI Nordics Discord community

GPT-SW3 consortium

AI SWEDEN

RISE

WASP | WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM
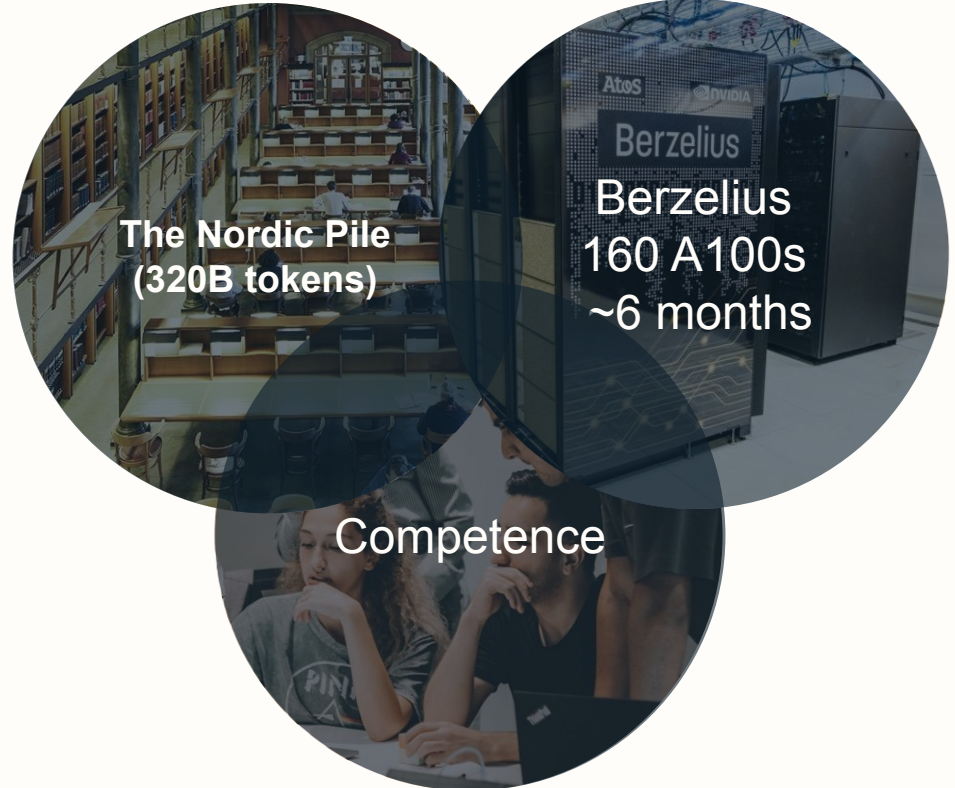
NVIDIA

AI Nordics Discord community

NSC

The Nordic Pile
(320B tokens)

Infrastructure

Competence

GPT-SW3 consortium

The Nordic Pile
(320B tokens)

Berzelius
160 A100s
~6 months

Competence

AI Nordics Discord community

Generative AI System Release Method Over Time (All Modalities)
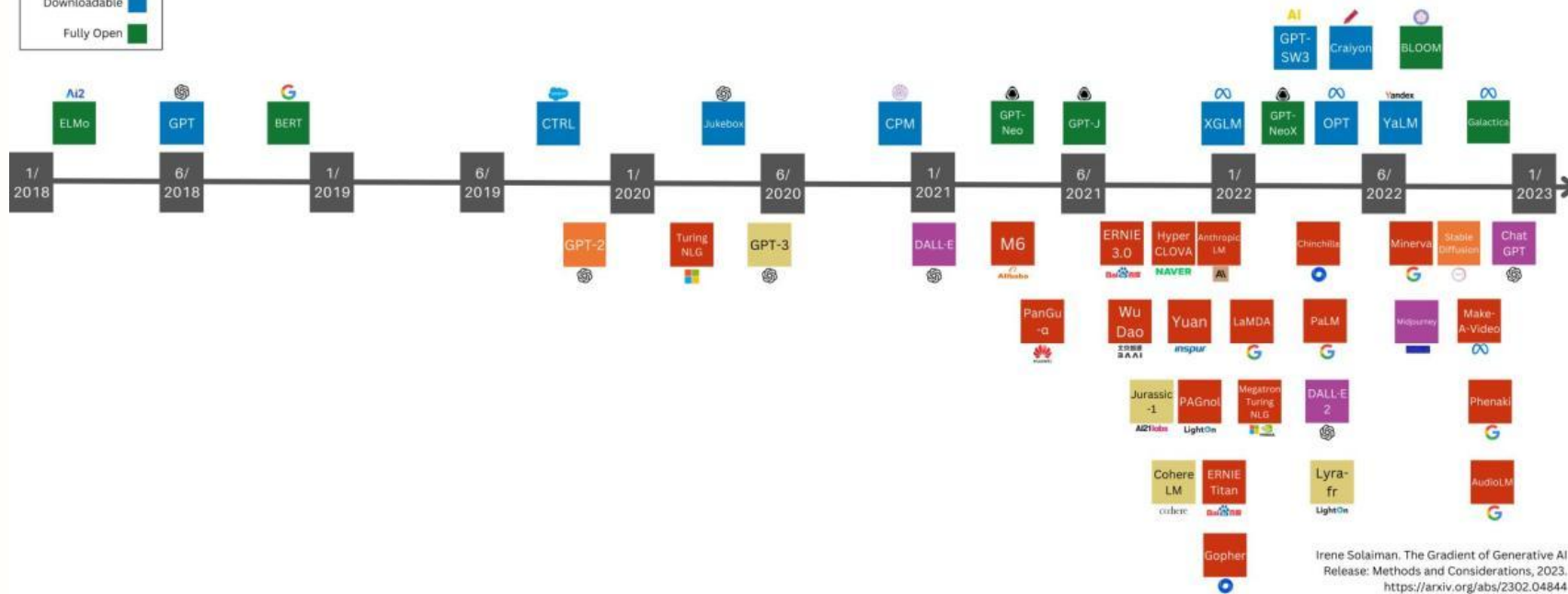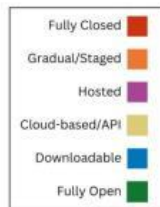
magnus.sahlgren@ai.se

Generative AI System Release Method Over Time (All Modalities)

magnus.sahlgren@ai.se

FinGPT
Viking
Poro
Ahma

NORA.LLM

GPT-SW3

magnus.sahlgren@ai.se
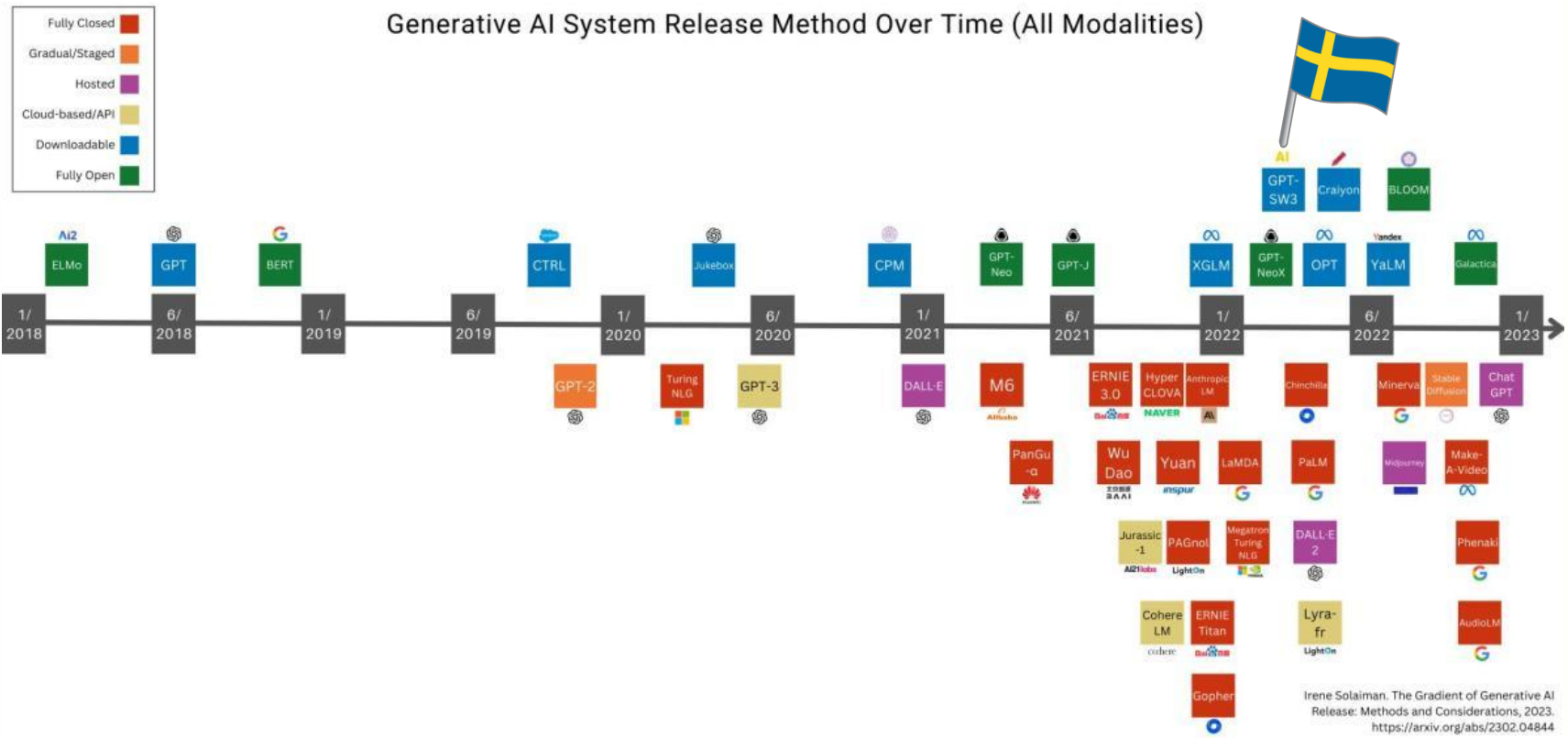
FinGPT
Viking
Poro
Ahma

NORA.LLM

GPT-SW3

Danish foundation models
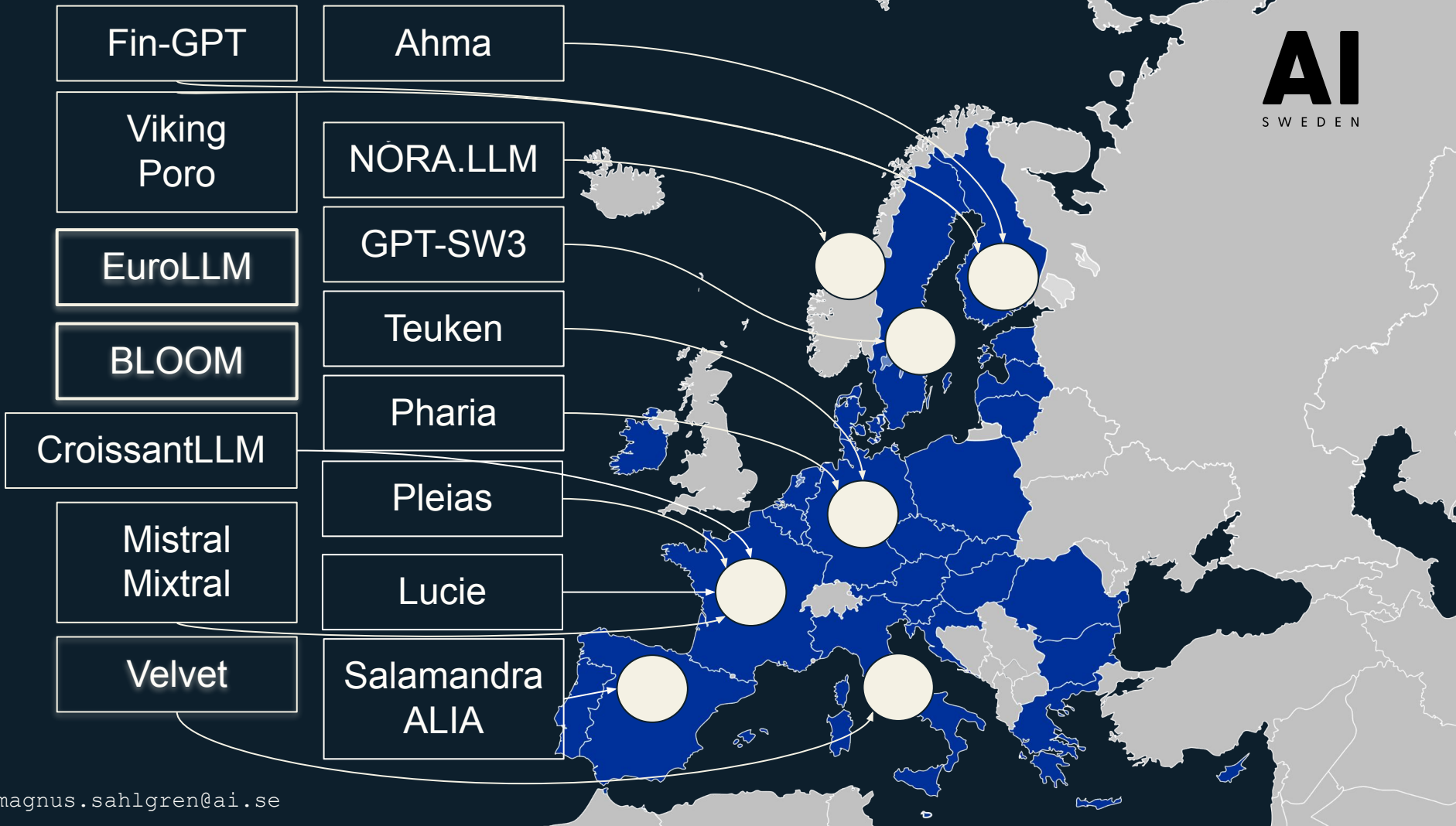
magnus.sahlgren@ai.se

Fin-GPT

Ahma

Viking
Poro

NORA.LLM

EuroLLM

GPT-SW3

BLOOM

Teuken

CroissantLLM

Pharia

Mistral
Mixtral

Pleias

Lucie

Velvet

Salamandra
ALIA

AI
SWEDEN

Open vs closed models

AI SWEDEN

Data compiled by Stella Biderman

Maxime Labonne

magnus.sahlgren@ai.se
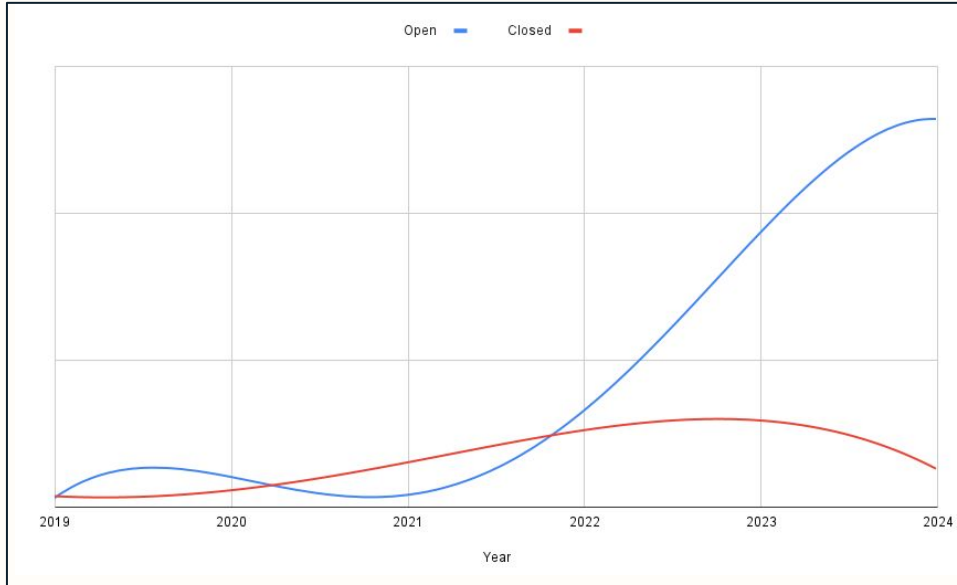
Open vs closed models
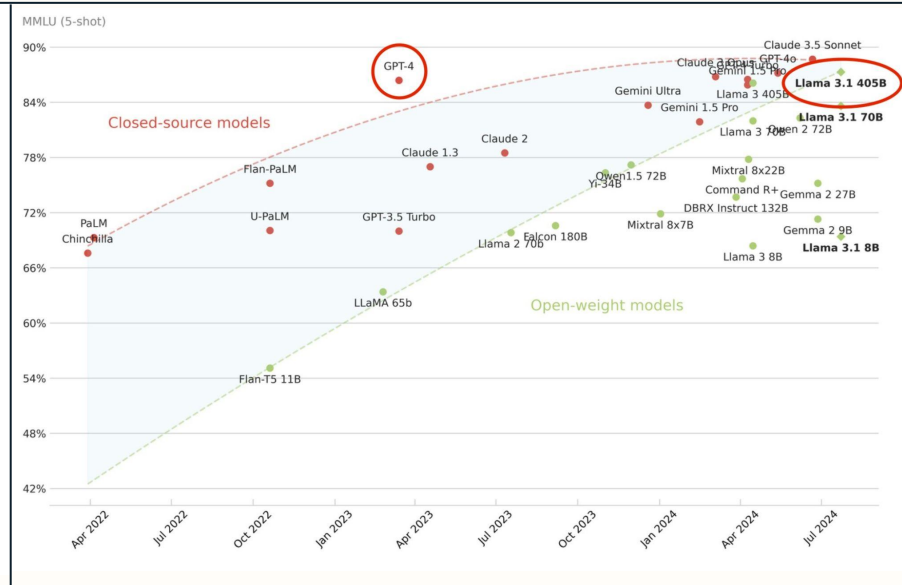
**Important distinction:
open source vs open weights**
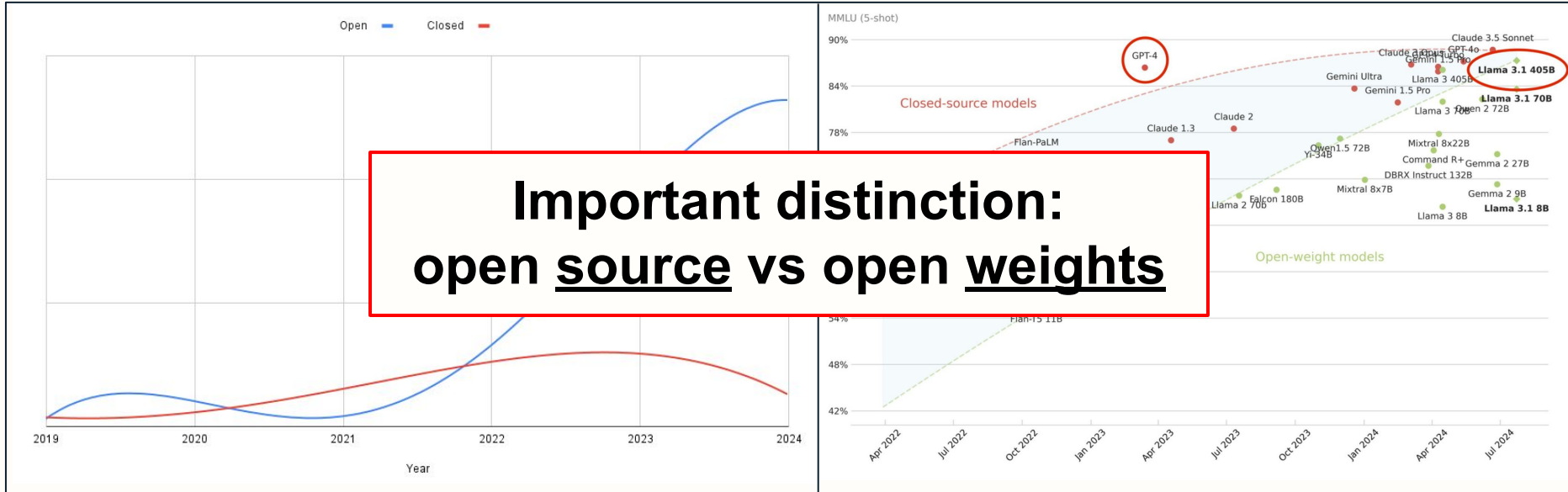
Data compiled by Stella Biderman

Maxime Labonne

Rank* (UB)	Rank (StyleCtrl)	Model	Arena Score	95% CI	Votes	Organization	License
1	1	Gemini-2.5-Pro-Preview-05-06	1448	+7/-12	3545	Google	Proprietary
2	1	o3-2025-04-16	1411	+11/-7	5844	OpenAI	Proprietary
2	2	ChatGPT-4o-latest (2025-03-26)	1408	+6/-5	10286	OpenAI	Proprietary
2	5	Grok-3-Preview-02-24	1402	+5/-4	14843	xAI	Proprietary
4	3	GPT-4.5-Preview	1398	+4/-5	15281	OpenAI	Proprietary
4	5	Gemini-2.5-Flash-Preview-04-17	1394	+9/-8	5393	Google	Proprietary
7	6	DeepSeek-V3-0324	1372	+6/-6	8213	DeepSeek	MIT
7	5	GPT-4.1-2025-04-14	1367	+6/-8	4523	OpenAI	Proprietary
8	7	DeepSeek-R1	1359	+4/-3	18034	DeepSeek	MIT
8	5	o4-mini-2025-04-16	1351	+9/-8	4526	OpenAI	Proprietary
9	15	Gemini-2.0-Flash-001	1355	+3/-4	24378	Google	Proprietary
10	7	o1-2024-12-17	1350	+3/-3	29042	OpenAI	Proprietary
10	14	Qwen3-235B-A22B	1342	+11/-9	2805	Alibaba	Apache 2.0

magnus.sahlgren@ai.se

Rank* (UB)	Rank (StyleCtrl)	Model	Arena Score	95% CI	Votes	Organization	License
1	1	Gemini-2.5-Pro-Preview-05-06	1448	+7/-12	3545	Google	Proprietary
2	1	o3-2025-04-16	1411	+11/-7	5844	OpenAI	Proprietary
2	2	ChatGPT-4o-latest (2025-03-26)	1408	+6/-5	10286	OpenAI	Proprietary
2	5	Grok-3-Preview-02-24	1402	+5/-4	14843	xAI	Proprietary
4	3	GPT-4.5-Preview	1398	+4/-5	15281	OpenAI	Proprietary
4	5	Gemini-2.5-Flash-Preview-04-17	1394	+9/-8	5393	Google	Proprietary
7	6	DeepSeek-V3-0324	1372	+6/-6	8213	DeepSeek	MIT
7	5	GPT-4.1-2025-04-14	1367	+6/-8	4523	OpenAI	Proprietary
8	7	DeepSeek-R1	1359	+4/-3	18034	DeepSeek	MIT
8	5	o4-mini-2025-04-16	1351	+9/-8	4526	OpenAI	Proprietary
9	15	Gemini-2.0-Flash-001	1355	+3/-4	24378	Google	Proprietary
10	7	o1-2024-12-17	1350	+3/-3	29042	OpenAI	Proprietary
10	14	Qwen3-235B-A22B	1342	+11/-9	2805	Alibaba	Apache 2.0

lmarena.ai

magnus.sahlgren@ai.se

The DeepSeek moment

DeepSeek claimed the model training took 2.8 million H800 GPU hours, which, at a cost of $2/GPU hour, comes out to a mere $5.6 million.

Why NVIDIA's share price dropped 17% after DeepSeek news

NVIDIA's stock suffered its largest single-day decline on Monday, falling 17% after Chinese startup DeepSeek unveiled a cost-effective AI model.

anton @abacaj · 12h
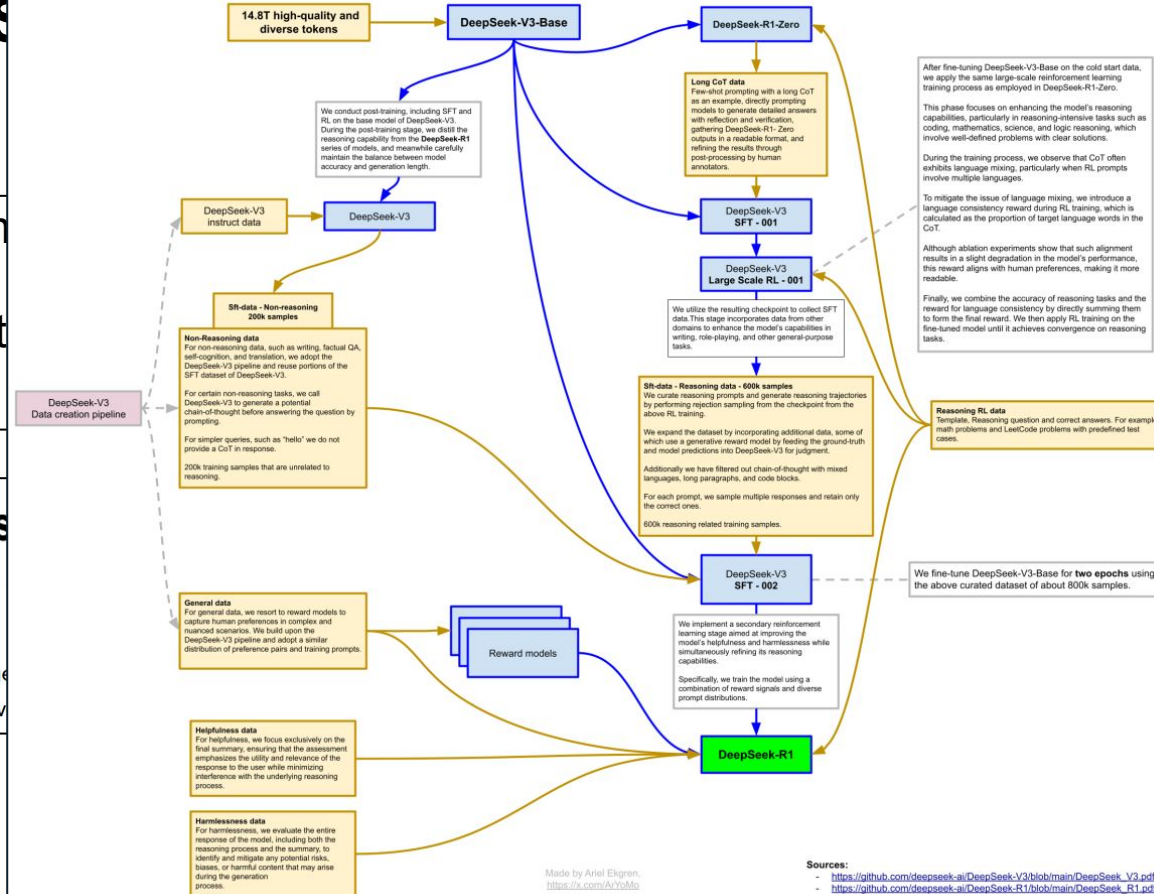How is Deepseek going to make money?
💬 541   🔁 171   ❤ 2.4K   📊 498K

Han Xiao @hxiao   Follow

deepseek's holding 幻方量化 is a quant company, many years already, super smart guys with top math background; happened to own a lot GPU for trading/mining purpose, and deepseek is their side project for squeezing those gpus

DeepSeek claim
took 2.8 million
which, at a cost
comes out to a

**Why NVIDIA's
dropped 17%
news**

NVIDIA's stock suffered its large
Chinese startup DeepSeek unv

DeepSeek-R1
datasets and models breakdown

	Model	Type	Rank ▲
1	gemini/gemini-1.5-pro-002 (few-shot, val)	📝	1.36
2	gemini/gemini-2.0-flash-001 (few-shot, val)	📝	1.37
3	gemini/gemini-2.0-flash-lite-001 (few-shot, val)	📝	1.43
4	gpt-4o-2024-05-13 (few-shot, val)	📝	1.46
5	o3-2025-04-16 (zero-shot, val)	🤔	1.57
6	gemini/gemini-1.5-flash-002 (few-shot, val)	📝	1.60
7	gpt-4.1-2025-04-14 (zero-shot, val)	📝	1.62
8	gemini/gemini-2.5-flash-preview-04-17 (zero-shot, val)	🤔	1.64
9	meta-llama/Llama-3.1-70B-Instruct (few-shot)	📝	1.67
10	o4-mini-2025-04-16 (zero-shot, val)	🤔	1.68
11	meta-llama/Meta-Llama-3-70B (few-shot)	🧠	1.68
12	gpt-4o-2024-05-13 (zero-shot, val)	📝	1.69
13	Qwen/Qwen2.5-72B-Instruct (few-shot)	📝	1.69
14	claude-3-5-sonnet-20241022 (zero-shot, val)	📝	1.69
15	meta-llama/Llama-3.3-70B-Instruct (few-shot)	📝	1.70

euroeval.com

magnus.sahlgren@ai.se

	Model	Type	Rank ▲
1	gemini/gemini-1.5-pro-002 (few-shot, val)	📝	1.36
2	gemini/gemini-2.0-flash-001 (few-shot, val)	📝	1.37
3	gemini/gemini-2.0-flash-lite-001 (few-shot, val)	📝	1.43
4	gpt-4o-2024-05-13 (few-shot, val)	📝	1.46
5	o3-2025-04-16 (zero-shot, val)	🤔	1.57
6	gemini/gemini-1.5-flash-002 (few-shot, val)	📝	1.60
7	gpt-4.1-2025-04-14 (zero-shot, val)	📝	1.62
8	gemini/gemini-2.5-flash-preview-04-17 (zero-shot, val)	🤔	1.64
9	meta-llama/Llama-3.1-70B-Instruct (few-shot)	📝	1.67
10	o4-mini-2025-04-16 (zero-shot, val)	🤔	1.68
11	meta-llama/Meta-Llama-3-70B (few-shot)	🧠	1.68
12	gpt-4o-2024-05-13 (zero-shot, val)	📝	1.69
13	Qwen/Qwen2.5-72B-Instruct (few-shot)	📝	1.69
14	claude-3-5-sonnet-20241022 (zero-shot, val)	📝	1.69
15	meta-llama/Llama-3.3-70B-Instruct (few-shot)	📝	1.70

euroeval.com

Best european model (Mistral) is on place 27

(Mistral has raised > $1B in investments)

magnus.sahlgren@ai.se

rank	model_name	total	hp	matematik_och_fysikprovet	läkare	tandläkare	audionom	apotekare	open_weights
1.	openai/o3-mini	0.85	1.9	0.77	0.89	0.80	0.82	0.96	❌
2.	anthropic/claude-3.7-sonnet	0.84	2.0	0.62	0.88	0.87	0.85	1.00	❌
3.	openai/gpt-4o	0.82	2.0	0.51	0.89	0.87	0.89	0.98	❌
4.	google/gemini-2.0-flash-001	0.82	2.0	0.63	0.87	0.81	0.81	0.96	❌
5.	meta-llama/llama-3.1-405b-instruct	0.81	2.0	0.55	0.91	0.76	0.82	0.96	✅
6.	deepseek/deepseek-V3	0.80	2.0	0.60	0.83	0.71	0.89	0.94	✅
7.	x-ai/grok-2-1212	0.80	2.0	0.50	0.88	0.80	0.82	0.98	❌
8.	meta-llama/llama-3.3-70b-instruct	0.78	1.9	0.53	0.82	0.77	0.82	0.94	✅
9.	qwen/qwen-max	0.77	1.7	0.60	0.84	0.66	0.78	0.92	❌
10.	cohere/command-a	0.72	1.8	0.48	0.78	0.70	0.72	0.89	✅
11.	mistralai/mistral-small-3.1-24b-instruct	0.71	1.6	0.50	0.78	0.61	0.76	0.91	✅
12.	google/gemma-3-27b-it	0.67	1.8	0.50	0.64	0.69	0.72	0.70	✅
13.	allenai/olmo-2-0325-32b-instruct	0.57	1.3	0.36	0.63	0.50	0.53	0.77	✅
14.	meta-llama/llama-3.2-3b-instruct	0.36	0.7	0.19	0.46	0.41	0.31	0.38	✅

magnus.sahlgren@ai.se

Why build national LLMs?

- Competence and innovation
- Transparency and representativity
- Resilience and sovereignty
- Democratisation

magnus.sahlgren@ai.se

Why build national LLMs?

- Competence and innovation
- Transparency and representativity
- Resilience and sovereignty
- Democratisation
- **Existing models still not perfect for smaller languages!**

magnus.sahlgren@ai.se

Adaptation of open models

magnus.sahlgren@ai.se

Adaptation of open models

magnus.sahlgren@ai.se
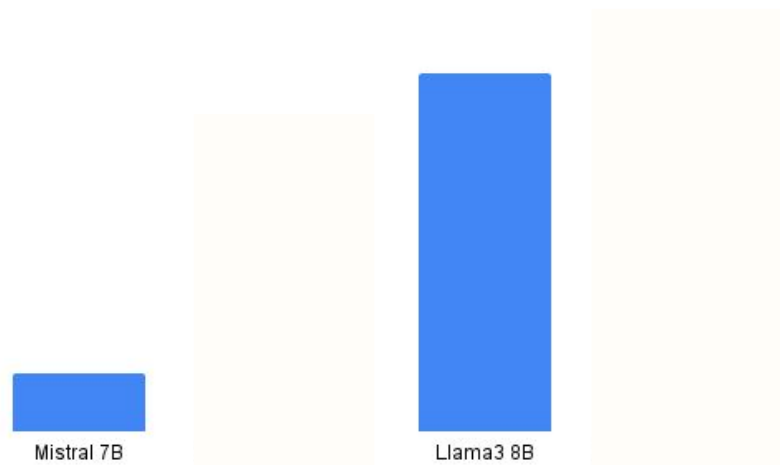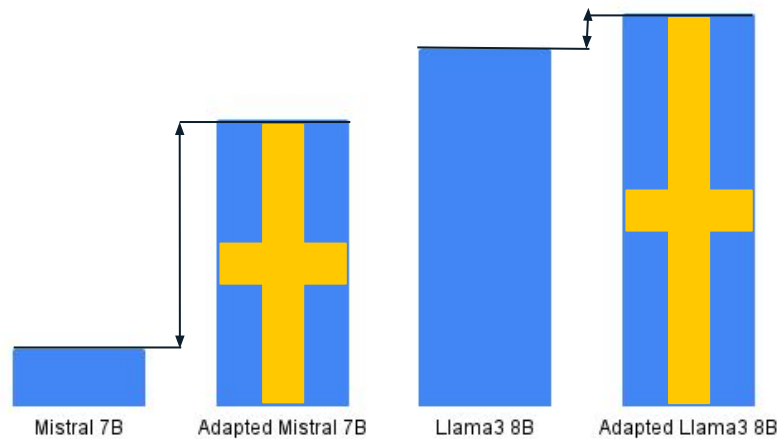
Competence

Language models
Deep learning
HPC

Data

Trillions of high-quality tokens
(We have to be compliant with
 European data protection policy)

AI
S W E D E N

Funding

Large-scale projects
Not research!
(The artefacts are the goal)

Infrastructure

1000s of high-end GPUs (H100s)
(with good interconnect)
Large-scale allocations!

magnus.sahlgren@ai.se

The flagship European initiative to compete with the US and China

A series of transparent, compliant and multilingual foundation models built in Europe for use in Europe

20M EUR 2025-2027

Universities and Research Organizations

AI SWEDEN

Charles University

AI SWEDEN

alt-edic el Da

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

ellis INSTITUTE TÜBINGEN

Fraunhofer IAIS

JÜLICH Forschungszentrum

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF HELSINKI

UNIVERSITAS OSLOENSIS · MDCCCXI ·

UNIVERSITY OF TURKU

Companies

ALEPH ALPHA

AMD SILO AI

ellamind

Lighton

prompsit

HPC Centres

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

CINECA

CSC

SURF

magnus.sahlgren@ai.se

huggingface.co/AI-Sweden-Models

AI AI-Sweden-Models/Llama-3-8B
Text Generation · Updated Jul 11 · ⬇ 105 · ♡ 17

AI AI-Sweden-Models/gpt-sw3-20b-instruct
Text Generation · Updated Jun 12 · ⬇ 4.09k · ♡ 12

AI AI-Sweden-Models/gpt-sw3-6.7b-v2-instruct
Text Generation · Updated Jan 29 · ⬇ 3.85k · ♡ 10

AI AI-Sweden-Models/gpt-sw3-40b
Text Generation · Updated Feb 20 · ⬇ 813 · ♡ 8

AI AI-Sweden-Models/Llama-3-8B-instruct
Text Generation · Updated Jul 11 · ⬇ 370 · ♡ 8

AI AI-Sweden-Models/gpt-sw3-6.7b-v2-translator
Text Generation · Updated Apr 2 · ⬇ 24 · ♡ 6

AI AI-Sweden-Models/gpt-sw3-6.7b-v2-instruct-4bit-gptq
Text Generation · Updated Dec 19, 2023 · ⬇ 123 · ♡ 5

AI AI-Sweden-Models/roberta-large-1160k
Fill-Mask · Updated Apr 20 · ⬇ 204 · ♡ 4

AI AI-Sweden-Models/gpt-sw3-126m
Text Generation · Updated Jan 29 · ⬇ 84.4k · ♡ 3

AI AI-Sweden-Models/gpt-sw3-1.3b
Text Generation · Updated Jan 29 · ⬇ 1.35k · ♡ 3

magnus.sahlgren@ai.se